HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

PATENT APPLICATION

ATTORNEY DOCKET NO. 200311046-1

IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): Ludmila Cherkasova et al.
Confirmation No.: 9816

Application No.: 10/601,992
Examiner: A.M. Bhatia

Filing Date: June 23, 2003
Group Art Unit: 2145

Title: COST-AWARE ADMISSION CONTROL FOR STREAMING MEDIA SERVER

Mail Stop Appeal Brief-Patents
Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

TRANSMITTAL OF APPEAL BRIEF

Transmitted herewith is the Appeal Brief in this application with respect to the Notice of Appeal filed on May 23, 2007.

The fee for filing this Appeal Brief is (37 CFR 1.17(c)) $500.00

(complete (a) or (b) as applicable)

The proceedings herein are for a patent application and the provisions of 37 CFR 1.136(a) apply.

☐ (a) Applicant petitions for an extension of time under 37 CFR 1.136 (fees: 37 CFR 1.17(a)-(d)) for the total number of months checked below:

☐ 1st Month $120 ☐ 2nd Month $450 ☐ 3rd Month $1020 ☐ 4th Month $1590

☐ The extension fee has already been filed in this application.

☒ (b) Applicant believes that no extension of time is required. However, this conditional petition is being made to provide for the possibility that applicant has inadvertently overlooked the need for a petition and fee for extension of time.

Please charge to Deposit Account 08-2025 the sum of $ 500 . At any time during the pendency of this application, please charge any fees required or credit any over payment to Deposit Account 08-2025 pursuant to 37 CFR 1.25. Additionally please charge any fees to Deposit Account 08-2025 under 37 CFR 1.16 through 1.21 inclusive, and any other sections in Title 37 of the Code of Federal Regulations that may regulate fees.

Respectfully submitted,

Ludmila Cherkasova et al.

By: Jody C. Bishop

Attorney/Agent for Applicant(s)

Reg No. : 44, 034

Date : July 19, 2007

Telephone : (214) 855-8007

I hereby certify that this document is being transmitted to the Patent and Trademark Office via eletronic filing.

Date of Transmission: July 19, 2007

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

Docket No.: 200311046-1
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Ludmila Cherkasova et al.

Application No.: 10/601,992 | Confirmation No.: 9816

Filed: June 23, 2003 | Art Unit: 2145

For: COST-AWARE ADMISSION CONTROL FOR STREAMING MEDIA SERVER | Examiner: A. M. Bhatia

APPEAL BRIEF

MS Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

As required under § 41.37(a), this brief is filed within two months of the Notice of Appeal filed in this case on May 23, 2007, and is in furtherance of said Notice of Appeal.

The fees required under § 41.20(b)(2) are dealt with in the accompanying TRANSMITTAL OF APPEAL BRIEF.

This brief contains items under the following headings as required by 37 C.F.R. § 41.37 and M.P.E.P. § 1205.2:

I. Real Party In Interest
II Related Appeals and Interferences
III. Status of Claims
IV. Status of Amendments
V. Summary of Claimed Subject Matter

I. REAL PARTY IN INTEREST

The real party in interest for this appeal is:

Hewlett-Packard Development Company, L.P., a Limited Partnership established under the laws of the State of Texas and having a principal place of business at 20555 S.H. 249, Houston, TX 77070, U.S.A. (hereinafter "HPDC"). HPDC is a Texas limited partnership and is a wholly-owned affiliate of Hewlett-Packard Company, a Delaware Corporation, headquartered in Palo Alto, CA. The general or managing partner of HPDC is HPQ Holdings, LLC.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

III. STATUS OF CLAIMS

A. Total Number of Claims in Application

There are 42 claims pending in application.

B. Current Status of Claims

1. Claims canceled: none
2. Claims withdrawn from consideration but not canceled: none
3. Claims pending: 1-11, 13-17 and 19-42
4. Claims allowed: None
5. Claims rejected: 1-11, 13-17 and 19-42

C. Claims On Appeal

The claims on appeal are claims 1-11, 13-17 and 19-42

IV. STATUS OF AMENDMENTS

A Final Office Action rejecting the claims of the present application was mailed February 23, 2007. In response, Applicant did not file an Amendment in response to the Final Office Action, but instead filed a Notice of Appeal, which this brief supports. Accordingly, the claims on appeal are those as rejected in the Final Office Action of February 23, 2007. A complete listing of the claims is provided in the Claims Appendix hereto.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The following provides a concise explanation of the subject matter defined in each of the separately argued claims involved in the appeal, referring to the specification by page and line number and to the drawings by reference characters, as required by 37 C.F.R. § 41.37(c)(1)(v). Each element of the claims is identified by a corresponding reference to the specification and drawings where applicable. It should be noted that the citation to passages in the specification and drawings for each claim element does not imply that the limitations from the specification and drawings should be read into the corresponding claim element.

According to one claimed embodiment, such as that of independent claim 1,a method for managing admission of requests (e.g., requests 107, 109, 111 of FIG. 1) to a streaming media server (e.g., server 101 of FIG. 1) comprises receiving (e.g., operational block 901 of FIG. 9) a new request for a streaming media file to be served by a streaming media server. The method further comprises performing a resource availability check for the streaming media server to determine whether the streaming media server has sufficient available resources to service the new request (e.g., operational block 904 of FIG. 9). The method further comprises performing a quality of service guarantee check for the streaming media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests (e.g., operational block 905 of FIG. 9). *See e.g.,* paragraphs 0009, 0029-0031, and 0118-0125 of the specification.

In certain embodiments, such as that of dependent claim 2, the resource availability check comprises using a segment-based memory model to determine whether at least a portion of the

requested streaming media file is in the streaming media server's memory, *see e.g.,* paragraphs 0031-0034 and 0052-0097 of the specification.

In certain embodiments, such as that of dependent claim 3, the method further comprises determining from the segment-based memory model a cost associated with serving the requested streaming media file from the streaming media server, *see e.g.,* paragraphs 0106-0132 and 0144.

In certain embodiments, such as that of dependent claim 4, the resource availability check comprises determining a cost associated with serving the requested streaming media file from the streaming media server, and in certain embodiments, such as that of dependent claim 5, the cost comprises a cost of serving the requested streaming media file either from memory or from disk, *see e.g.,* paragraphs 0106-0132 and 0144.

In certain embodiments, such as that of dependent claim 10, the method further comprises:

if determined that the streaming media server does not have sufficient available resources to service the new request or determined that acceptance of the new request will violate, at any point in the future, said desired quality of service provided by the streaming media server for any previously accepted requests, then rejecting the new request for service by the streaming media server, *see e.g.,* operational block 906 of FIG. 9 and paragraph 0122 of the specification.

In certain embodiments, such as that of dependent claim 41, the performing said quality of service guarantee check comprises performing said quality of service guarantee check even when determined by said resource availability check that the streaming media server has sufficient available resources to service the new request, *see e.g.,* operational blocks 904-905 of FIG. 9 and paragraphs 0121-0122 of the specification.

According to another claimed embodiment, such as that of independent claim 11, a method for managing admission of requests (e.g., requests 107, 109, 111 of FIG. 1) to a media server (e.g., server 101 of FIG. 1) comprises receiving a new request for a streaming file to be served by a media server (e.g., operational block 901 of FIG. 9). The method further comprises determining a cost to the media server for serving the requested streaming file, wherein the cost

corresponds to the media server's resources to be consumed in serving the requested streaming file (e.g., operational block 903 of FIG. 9). The determining said cost comprises determining a segment-based memory model that identifies content of the media server's memory as of a time that the new request is received, and using the segment-based memory model to determine whether at least a portion of the requested streaming file is in the media server's memory, *see e.g.*, paragraphs 0031-0034, 0052-0097, 0106-0132 and 0144 of the specification. The method further comprises determining, based at least in part on the cost, whether to admit the new request for service by the media server (e.g., operational blocks 904-907 of FIG. 9), and *see e.g.*, paragraphs 0010, 0029-0031, and 0118-0125 of the specification.

In certain embodiments, such as that of dependent claim 13, the cost comprises a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the media server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the media server's memory, *see e.g.*, paragraphs 0106-0132 and 0144.

In certain embodiments, such as that of dependent claim 16, determining whether to admit the new request for service by the media server further comprises performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests, *see e.g.*, operational block 905 of FIG. 9 and paragraphs 0121-0122 of the specification.

According to another claimed embodiment, such as that of independent claim 17, a system comprises a server (e.g., server 101 of FIG. 1) having a memory, wherein said server is operable to serve at least one streaming file to clients (e.g., clients 104, 105, and 106 of FIG. 1) communicatively coupled thereto. The system further comprises an admission controller operable to receive a new request (e.g., new requests 107, 109, and 11 of FIG. 1) for a streaming file to be served by said server, determine a cost to the server for serving the requested streaming file, wherein the cost corresponds to the server's resources to be consumed in serving the requested streaming file, and determine, based at least in part on the cost, whether to admit the

new request for service by the server (*see e.g.*, paragraphs 0011, 0029-0033, 0106-0132 and 0144). The admission controller is further operable to determine a segment-based memory model that identifies content of the server's memory as of a time that the new request is received, and said admission controller is operable to use the segment-based memory model to determine whether at least a portion of the requested streaming file is in the server's memory, *see e.g.*, paragraphs 0031-0034, 0052-0097, 0106-0132 and 0144 of the specification.

In certain embodiments, such as that of dependent claim 19, the cost comprises a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the server's memory, *see e.g.*, paragraphs 0106-0132 and 0144.

In certain embodiments, such as that of dependent claim 22, the admission controller is further operable to perform quality of service guarantee check for the server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the server for any previously accepted requests, *see e.g.*, operational block 905 of FIG. 9 and paragraphs 0121-0122 of the specification.

In certain embodiments, such as that of dependent claim 42, the admission controller determines the cost to the server for serving the requested streaming file based at least in part on the determined segment-based memory model, *see e.g.*, paragraphs 0031-0034, 0052-0097, 0106-0132 and 0144 of the specification.

According to another claimed embodiment, such as that of independent claim 23, a method comprises receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server (e.g., operational block 801 of FIG. 8, and operational block 901 of FIG. 9). The method further comprises creating a segment-based model of the media server's memory as of time T_{cur} (e.g., operational blocks 802-803 of FIG. 8, and operational block 902 of FIG. 9); and based at least in part on the segment-based model of the media server's memory, determining whether to accept the received request for service by the media server (e.g., operational block

804 of FIG. 8 and operational blocks 903-907 of FIG. 9), *see* paragraphs 0012, 0029-0031, 0082-0105, and 0118-0125 of the specification.

In certain embodiments, such as that of dependent claim 24, the segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment, *see e.g.* paragraphs 0062-0074 of the specification.

In certain embodiments, such as that of dependent claim 28, the determining whether to accept the received request for service by the media server comprises performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests, *see e.g.,* operational block 905 of FIG. 9 and paragraphs 0121-0122 of the specification.

According to another claimed embodiment, such as that of independent claim 29, computer-executable software stored to a computer-readable medium comprises code for creating a segment-based model of a media server's memory (e.g., operational blocks 802-803 of FIG. 8, and operational block 902 of FIG. 9). The computer-executable software code further comprises code for determining whether to serve a requested streaming file from the media server based at least in part on the segment-based model of the media server's memory (e.g., operational block 804 of FIG. 8 and operational blocks 903-907 of FIG. 9), *see* paragraphs 0012, 0029-0031, 0082-0105, and 0118-0125 of the specification.

In certain embodiments, such as that of dependent claim 33, the segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment, *see e.g.* paragraphs 0062-0074 of the specification.

In certain embodiments, such as that of dependent claim 36,the code for determining whether to serve a requested streaming file from the media server comprises code for performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests, *see e.g.,* operational block 905 of FIG. 9 and paragraphs 0121-0122 of the specification.

According to another claimed embodiment, such as that of independent claim 37, a cost-aware admission control system comprises means (e.g., an admission controller and/or software stored to computer-readable medium and being executed by a computer, *see e.g.,* paragraphs 0011 and 0157-0163) for receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server (e.g., operational block 801 of FIG. 8, and operational block 901 of FIG. 9). The system further comprises means (e.g., an admission controller and/or software stored to computer-readable medium and being executed by a computer, *see e.g.,* paragraphs 0011 and 0157-0163) for creating a segment-based model of the media server's memory as of time T_{cur} (e.g., operational blocks 802-803 of FIG. 8, and operational block 902 of FIG. 9); and means (e.g., an admission controller and/or software stored to computer-readable medium and being executed by a computer, *see e.g.,* paragraphs 0011 and 0157-0163) for determining, based at least in part on the segment-based model of the media server's memory, whether to accept the received request for service by the media server (e.g., operational block 804 of FIG. 8 and operational blocks 903-907 of FIG. 9), *see* paragraphs 0012, 0029-0031, 0082-0105, and 0118-0125 of the specification.

In certain embodiments, such as that of dependent claim 38, the segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment, *see e.g.* paragraphs 0062-0074 of the specification.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A. Claims 29-36 are rejected under 35 U.S.C. § 101 as being directed to non-statutory matter.

B. Claims 1-2, 6-8, 23-25, 29-33, and 37-39 are rejected under 35 U.S.C. § 102(b) as being anticipated by U.S. Patent No. 5,991,811 to Ueno et al. (hereinafter "*Ueno*").

C. Claims 3-5, 9-11, 13-17, 19-22, 26-28, 35-36, and 40-42 are rejected under 35 U.S.C. § 103(a) as being unpatentable over *Ueno* in view of U.S. Patent No. 6,910,024 to Krishnamurthy et al. (hereinafter "*Krishnamurthy*").

VII. ARGUMENT

Appellant respectfully traverses the outstanding rejections of the pending claims, and requests that the Board reverse the outstanding rejections in light of the remarks contained herein. The claims do not stand or fall together. Instead, Appellant presents separate arguments for various claims. Each of these arguments is separately argued below and presented with separate headings and sub-heading as required by 37 C.F.R. § 41.37(c)(1)(vii).

A. Rejections Under 35 U.S.C. § 101

Claims 29-36 are rejected under 35 U.S.C. § 101 as being directed to non-statutory matter. The Final Office Action appears to maintain that these claims are directed to "a signal encoded with functional descriptive material", *see* page 3 of the Final Office Action. Appellant respectfully disagrees and submits that the claims are directed to proper statutory subject matter under 35 U.S.C. §101, as discussed below.

Appellant respectfully asserts that the rejected language of claims 29-36, which recites "Computer-executable software stored to a computer-readable medium", is not merely directed to a signal. Instead, claims 29-36 are directed to a well-accepted type of claim, *see e.g., In re Beauregard*, 53 F.3d 1583 (Fed. Cir. 1995). M.P.E.P. §2106 further explains that "When functional descriptive material is recorded on some computer-readable medium it becomes structurally and functionally interrelated to the medium and will be statutory in most cases since use of technology permits the function of the descriptive material to be realized."

In support of the rejection, the Examiner directs Applicant's attention to the Official Gazette Notices dated 22 November 2005, under Guidelines for Subject Matter Eligibility, *see* page 5 of the Final Office Action. However, such Guidelines explain:

> Data structures not claimed as embodied in computer-readable media are descriptive material per se and are not statutory because they are not capable of causing functional change in the computer. See, e.g., Warmerdam, 33 F.3d at 1361, 31 USPQ2d at 1760 (claim to a data structure per se held nonstatutory). Such claimed data structures do not define any structural and functional interrelationships between the data structure and other claimed aspects of the

> invention which permit the data structure's functionality to be realized. In contrast, a claimed computer-readable medium encoded with a data structure defines structural and functional interrelationships between the data structure and the computer software and hardware components which permit the data structure's functionality to be realized, and is thus statutory. (Emphasis added).
>
> Similarly, computer programs claimed as computer listings per se, i.e., the descriptions or expressions of the programs, are not physical "things." They are neither computer components nor statutory processes, as they are not "acts" being performed. Such claimed computer programs do not define any structural and functional interrelationships between the computer program and other claimed elements of a computer which permit the computer program's functionality to be realized. In contrast, a claimed computer-readable medium encoded with a computer program is a computer element which defines structural and functional interrelationships between the computer program and the rest of the computer which permit the computer program's functionality to be realized, and is thus statutory. (Emphasis added). See Lowry, 32 F.3d at 1583-84, 32 USPQ2d at 1035.

Thus, this makes clear that when, as in the present case, claims are directed to computer-executable software code that is stored to a computer-readable medium, such claims are statutory. Accordingly, claims 29-36 are directed to proper statutory subject matter, and therefore Appellant requests that this rejection of these claims be overturned.

B. Rejections Under 35 U.S.C. § 102(b) Over *Ueno*

Claims 1-2, 6-8, 23-25, 29-33, and 37-39 are rejected under 35 U.S.C. § 102(b) as being anticipated by *Ueno*. Appellant respectfully traverses these rejections below.

In order to anticipate a claim under 35 U.S.C. § 102, a reference must teach every element of the claim. *See* M.P.E.P. § 2131. *Ueno* does not teach every element of claims 1-2, 6-8, 23-25, 29-33, and 37-39, as discussed below, and therefore the rejection of these claims should be overturned.

<u>**Independent Claim 1 and Dependent Claims 6-8**</u>

Independent claim 1 recites:

> A method for managing admission of requests to a streaming media server, the method comprising:
> receiving a new request for a streaming media file to be served by a streaming media server;
> performing a resource availability check for the streaming media server to determine whether the streaming media server has sufficient available resources to service the new request; and
> <u>performing a quality of service guarantee check for the streaming media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests</u>. (Emphasis added).

Ueno fails to teach at least the above-emphasized element of claim 1. That is, as discussed below, *Ueno* fails to teach performing a quality of service guarantee check for a streaming media server to determine whether acceptance of a new request will violate, <u>at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests</u>.

Ueno proposes a "method for decreasing communication costs particularly when the service of storage type data such as VOD [video on demand] is offered in real time." Col. 7, lines 61-63 of *Ueno*. *Ueno* proposes a technique in which an information source (e.g., a movie) is transmitted from a server to a client. As shown in Fig. 2 of *Ueno*, the transmission technique

communicates the information source (stored to storage unit 201) from a server 200 to a headend 212, which in turn communicates the information source to set-top unit 211 of a client.

The transmission technique for sending the information source from storage unit 201 of server 200 to headend 212 uses two different classes of communication for such transmission. First, the server 200 begins sending to the headend 212 (via port 201a) the information source from the information source's beginning toward its end (i.e., "forward playing") using an expensive class of communication that guarantees real time communication, such as CBR. Also, the server 200 uses a less expensive class of communication, such as a best effort class (e.g., ABR or UBR), to begin transmitting to the headend 212 (via port 201b) the information source from the information source's end toward its beginning (i.e., "reverse playing"). The reverse playing portion of the information source that is transmitted using a best effort class of communication is stored at storage 207 of headend 212 until the forward playing portion reaches such stored portion.

Once the forward playing portion of the transmission (e.g., being communicated using guaranteed real-time communication class via port 201a) reaches the portion of the information source that is stored in storage 207, the communication from server 200 to headend 212 can be terminated and switch 209 of headend 212 causes headend 212 to begin communicating the remainder of the information source from storage unit 207 to set-top unit 211. Thus, a portion of the communication from server 200 to headend 212 can be via a less expensive class of communication (e.g., best effort classes ABR or UBR), thereby reducing the overall cost of the communication between the server 200 and headend 212.

Ueno explains this transmission technique at col. 7, line 64- col. 8, line 29 as follows:

> A server system, which has received a demand for the offer of image, begins the offer of information to a receiving apparatus, which has transmitted the demand, in real time in the forward direction from the beginning of an information source such as a movie. As the quality in this case, the real-time characteristic and the low cell-discarding characteristic are required. Therefore, a class of a sufficiently guaranteed quality, such as the CBR, is used. On the other hand, at the same time as the offer of call, an information, which is not required to

> be offered in real time, is transmitted to the same receiving apparatus, for example, in the retrospective direction from the end of the information source, using a best effort communication class, such as the ABR and the UBR, as a second communication line, which is inexpensive although it is in non real time.
>
> The receiving apparatus transfers the real-time information transferred through the CBR, to the downward thereof as it is. On the other hand, the non-real-time information transferred through the best effort is being stored in a storage device, and the call for the high-quality class for real-time is released at the point of time when the non-real-time information reaches the same information as the real-time information. The information after that point of time is not required to be transferred through the CBR since it has been already stored in the storage device.
>
> That is, if it is assumed that one information source is transferred, the portion of information transferred through the second-class communication line is unnecessary as the period of time for requesting to ensure the call of a high-quality class, and the transfer can be carried out at a low communication cost by that portion. In particular, if the second communication line is the best effort class, the transfer can be extremely effectively carried out.

Ueno describes at col. 18, lines 18-54, for example, that there are limits on the number of simultaneous user accesses of the same video source. For instance, col. 18, lines 35-51 explains:

> Since there are upper limits with respect to the number of simultaneous accesses to the same video source and the number of users to which services are able to be offered at the same time for each server, a new demand for service is not able to be accepted when they have already reached the upper limits.... Since there are upper limits in the bands of transmission lines and the buffer capacities of exchange nodes, a new channel is not able to be established when they have already reached the upper limits.

This appears to suggest that a resource availability check may be performed to determine whether sufficient resources (e.g., bands of transmission lines) are available to service a new request. However, *Ueno* fails to teach performing a quality of service guarantee check for a streaming media server <u>to determine whether acceptance of a new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests</u>, as is further recited by claim 1. *Ueno* appears to merely evaluate whether a pre-set upper limit on the resources (e.g., transmission lines) is reached in determining whether to accept a new request. *Ueno* does not evaluate whether acceptance of the new request

will violate a desired quality of service for any previously accepted requests. Indeed, *Ueno* appears to suggest that the upper limits are set such that if not reached it can be assumed that acceptance of a new request will not impact already accepted requests. In any case, *Ueno* simply provides no teaching whatsoever of performing a quality of service guarantee check to determine if acceptance of a new request will cause violation of a desired quality of service for any previously accepted requests. Again, an analysis of the impact on the quality of service for already accepted requests is not performed in determining whether to accept a new request in *Ueno*.

The Final Office Action cites to Col. 16, line 49 – col. 17, line 26 of *Ueno* as teaching this element of claim 1, *see* page 6 of the Final Office Action. The cited portion of *Ueno* provides:

> In this preferred embodiment, a real-time data is transmitted via a single line 713, which offers the ABR service class. In the ABR, parameters such as a guaranteed minimum transmission rate (which will be hereinafter referred to as "MinR") and a peak rate (which will be hereinafter referred to as "PeakR") are reported to establish communication. That is, if there is room in bands after ensuring at least the bands of the MinR, the transmission rate up to the PeakR is able to be obtained. Therefore, it is assumed that the transmission rate for reproducing a transmitted real-time data in real time is Rr. If the report and transmission are performed so that Rr≤MinR, the arrival of data at the receiving side is not delayed from the required time. In this method, the real-time data is transmitted to the input port of a FIFO 707. The read-out rate at the output of the FIFO 707 is determined by the principle described in FIG. 5, and it is read out at a rate of Rr to be inputted to a decoder 712 to be reproduced in real time. In the FIFO 707, integrated values of actual transmission rates Rr are stored. When the transmission of all the data from the transmitting side is finally completed, this data of the integrated values corresponds to the portion which has been able to be transmitted by the room for bands in the network, so that it is possible to naturally transmit that portion at a lower cost than that of the transmission via the CBR. The quantity of data stored in the FIFO 707 is monitored by stored-data quantity monitoring means 719. On the basis of this quantity of stored data, the scope of transmission rate exceeding the MinR on the transmitting side is controlled by a control signal by means of communication control means 718 and 717, in order to prevent the FIFO 707 from overflowing when the capacity of the FIFO 707 is less than the capacity (=T(1-MinR/PeakR), T: quantity of all data) for storing all the maximum values of the finitely transmitted real-time data. The most simple method is a method for controlling the transmission rate on the transmitting side

so as to be the MinR after the FIFO 707 becomes a certain value Th1. This state is shown in FIG. 8. The Th1 is derived from the following formulae:

$$F = N / MinR * (MinR - Rr) + Th1$$

$$N = T - Rr * t - Th1$$

wherein F: Capacity of FIFO, t: Reproduction Time up to the Present, and N: Quantity of Data remaining at Transmitting Side.

This portion of *Ueno* appears to describe a transmission technique using a best effort class of communication (e.g., ABR) to communicate data according to certain transmission rate parameters (e.g., set between MinR and PeakR). The receiving set-top unit stores the received data to a FIFO 707, and the transmission rate of the data to the set-top unit is controlled such that the FIFO 707 does not overflow. Thus, this merely describes controlling transmission rate of data from a server to a set-top unit to enable the FIFO on the set-top unit to handle the received data without overflowing. This in no way teaches performing a quality of service guarantee check to determine if acceptance of a new request will cause violation of a desired quality of service for any previously accepted requests. Instead, this portion of *Ueno* merely addresses controlling the transmission rate of a given stream so that the stream can be handled by the FIFO of a set-top unit, <u>without any consideration of the quality of service of other previously accepted requests</u>.

Again, an analysis of the impact on the quality of service for already accepted requests is not performed in determining whether to accept a new request in *Ueno*. In response to the above arguments, the Final Office Action asserts that "because the service guarantee is provided continuously in Ueno it also inherent would include any new request." Page 2 of the Final Office Action. First, this statement by the Examiner fails to even allege that *Ueno* analyzes the impact on the quality of service for <u>already accepted requests</u> in determining whether to accept a new request. Instead, it merely asserts that the service guarantee of *Ueno* includes the new request itself (without any consideration for already accepted requests). Additionally, this unsupported statement alone is insufficient to establish inherency. In order to properly establish that an element is inherently included within an applied reference, "the Examiner must provide a

basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic necessarily flows from the teachings of the applied prior art," M.P.E.P. § 2112, citing *Ex parte Levy*, 17 U.S.P.Q.2d 1461, 1464 (Bd. Pat. App. & Inter. 1990) (emphasis original). The Examiner has failed to provide such a basis in fact and/or technical reasoning. Indeed, as discussed below, the alleged inherent characteristic does not necessarily flow from the teaching of *Ueno*.

As discussed above, it appears that Ueno merely analyzes whether to accept a new request without any consideration for the impact of such an acceptance on previously accepted requests. That is, *Ueno* appears to merely evaluate whether a pre-set upper limit on the resources (e.g., transmission lines) is reached in determining whether to accept a new request. *Ueno* does not evaluate whether acceptance of the new request will violate a desired quality of service for any previously accepted requests. Indeed, *Ueno* appears to suggest that the upper limits are set such that if the limit is not reached then it is assumed that acceptance of a new request will not impact already accepted requests. In any event, *Ueno* does not perform a quality of service guarantee check for the streaming media server to determine whether acceptance of a new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests, but instead *Ueno* merely evaluates whether a pre-set upper limit on the resources is reached when determining whether to accept a new request (without any consideration of future impact of a newly accepted request on previously accepted requests).

In view of the above, *Ueno* fails to teach all elements of claim 1, and therefore the rejection of claim 1 should be overturned.

Claims 6-8 each depend either directly or indirectly from independent claim 1, and are thus likewise believed to be allowable at least based on their dependency from claim 1 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claims 6-8 also be overturned.

Dependent Claim 2

Dependent claim 2 depends from claim 1, and thus inherits all of the limitations of claim 1 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 2 is allowable at least because of its dependence from claim 1 for the reasons discussed above.

Further, claim 2 recites "wherein said resource availability check comprises: using a segment-based memory model to determine whether at least a portion of the requested streaming media file is in the streaming media server's memory." *Ueno* fails to teach this further element of claim 2. That is, as discussed below, *Ueno* fails to teach using a segment-based memory model, as recited by claim 2. *Ueno* provides no teaching of using a segment-based memory model. *Ueno* does not appear to teach any model of a media server's memory, and certainly not a segment-based model of such memory.

The Final Office Action (at page 6 thereof) cites to col. 11, lines 10-25, col. 8, lines 1-11, and col. 16, line 49 - col. 17, line 26 of *Ueno* as teaching such use of a segment-based memory. As discussed below, the cited portions of *Ueno* provide no such teaching.

First, as discussed above with claim 1, col. 16, line 49 – col. 17, line 26 of *Ueno* merely describes controlling transmission rate of data from a server to a set-top unit to enable the FIFO on the set-top unit to handle the received data without overflowing. This in no way teaches use of a segment-based model to determine whether at least a portion of the requested streaming media file is in the streaming media server's memory. Indeed, this fails to teach any model of a media server's memory whatsoever.

Further, col. 11, lines 10-25 of *Ueno* provides:

> While it has been described that the source is transmitted from the output port 201b in order from the end, it is not always required to be transmitted in order from the very end in the case of a long source, but the whole source may be divided into a number of parts to perform the aforementioned transmission with respect to each part. In this case, the switching of the stream, the setting of connection and the disconnection as set forth above are repeated by the number of the division. In the aforementioned preferred embodiment, while the data stored

> in the storage unit has been transmitted at the same time as the transmission of the data reproduced in real time, it is considered that the data stored therein are not always transmitted at the same time so that they have been previously stored. This is the case that, for example, the storage unit is commonly used by a plurality of clients, and when a client accesses the program watched by a previous client, the data stored for the previous client have been already stored in the storage unit to be reused.

As can be clearly seen, this does not provide any teaching whatsoever of using a model of a media server's memory, and certainly fails to teach using a segment-based memory model as recited by claim 2.

Further, col. 8, lines 1-11 of *Ueno* provides:

> As the quality in this case, the real-time characteristic and the low cell-discarding characteristic are required. Therefore, a class of a sufficiently guaranteed quality, such as the CBR, is used. On the other hand, at the same time as the offer of call, an information, which is not required to be offered in real time, is transmitted to the same receiving apparatus, for example, in the retrospective direction from the end of the information source, using a best effort communication class, such as the ABR and the UBR, as a second communication line, which is inexpensive although it is in non real time.

As can be clearly seen, this further portion of *Ueno* also does not provide any teaching whatsoever of using a segment-based memory model as recited by claim 2. Instead, this portion of *Ueno* merely describes that different classes of communication, such as guaranteed quality class (e.g., CBR) and best effort class (e.g., ABR and UBR) can be used for communicating data.

In view of the above, *Ueno* fails to teach this further element of claim 2, and therefore the rejection of claim 2 should also be overturned.

<u>Independent Claim 23 and Dependent Claim 25</u>

Independent claim 23 recites:

> A method comprising:
> receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server;

> creating a segment-based model of the media server's memory as of time T_{cur}; and
>
> based at least in part on the segment-based model of the media server's memory, determining whether to accept the received request for service by the media server. (Emphasis added).

Ueno fails to teach at least the above-emphasized elements of claim 23. That is, as discussed below, *Ueno* fails to teach creating a segment-based model of a media server's memory as of the time T_{cur} at which a new request for a streaming media file is received by the media server. Further, *Ueno* fails to determine whether to accept the new request based at least in part on such a segment-based model.

Ueno provides no teaching of creating a segment-based model of a media server's memory as of the time T_{cur} at which a new request for a streaming media file is received by the media server. *Ueno* does not appear to teach any model of a media server's memory, and certainly not a segment-based model of such memory. The Final Office Action (at page 7 thereof) cites to col. 11, lines 10-25, col. 8, lines 1-11, and col. 16, line 49 - col. 17, line 26 of *Ueno* as teaching creating such a segment-based model of a media server's memory. As discussed below, the cited portions of *Ueno* provide no such teaching.

First, as discussed above with claim 1, col. 16, line 49 – col. 17, line 26 of *Ueno* merely describes controlling transmission rate of data from a server to a set-top unit to enable the FIFO on the set-top unit to handle the received data without overflowing. This in no way teaches creating a segment-based model of a media server's memory. Indeed, this fails to teach creating any model of a media server's memory whatsoever.

Further, col. 11, lines 10-25 of *Ueno* provides:

> While it has been described that the source is transmitted from the output port 201b in order from the end, it is not always required to be transmitted in order from the very end in the case of a long source, but the whole source may be divided into a number of parts to perform the aforementioned transmission with respect to each part. In this case, the switching of the stream, the setting of connection and the disconnection as set forth above are repeated by the number of the division. In the aforementioned preferred embodiment, while the data stored

> in the storage unit has been transmitted at the same time as the transmission of the data reproduced in real time, it is considered that the data stored therein are not always transmitted at the same time so that they have been previously stored. This is the case that, for example, the storage unit is commonly used by a plurality of clients, and when a client accesses the program watched by a previous client, the data stored for the previous client have been already stored in the storage unit to be reused.

As can be clearly seen, this does not provide any teaching whatsoever of creating a model of a media server's memory, and certainly fails to teach creating a segment-based model of the media server's memory as recited by claim 23.

Further, col. 8, lines 1-11 of *Ueno* provides:

> As the quality in this case, the real-time characteristic and the low cell-discarding characteristic are required. Therefore, a class of a sufficiently guaranteed quality, such as the CBR, is used. On the other hand, at the same time as the offer of call, an information, which is not required to be offered in real time, is transmitted to the same receiving apparatus, for example, in the retrospective direction from the end of the information source, using a best effort communication class, such as the ABR and the UBR, as a second communication line, which is inexpensive although it is in non real time.

As can be clearly seen, this further portion of *Ueno* also does not provide any teaching whatsoever of creating a model of a media server's memory, and certainly fails to teach creating a segment-based model of the media server's memory as recited by claim 23. Instead, this portion of *Ueno* merely describes that different classes of communication, such as guaranteed quality class (e.g., CBR) and best effort class (e.g., ABR and UBR) can be used for communicating data.

Further, because no such segment-based model of the media server's memory is created in *Ueno*, *Ueno* fails to teach determining whether to accept a new request based at least in part on such a segment-based model.

In view of the above, *Ueno* fails to teach all elements of claim 23, and therefore the rejection of claim 23 should be overturned.

Claim 25 depends from independent claim 23, and is thus likewise believed to be allowable at least based on its dependency from claim 23 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claim 25 also be overturned.

Dependent Claim 24

Dependent claim 24 depends from claim 23, and thus inherits all of the limitations of claim 23 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 24 is allowable at least because of its dependence from claim 23 for the reasons discussed above.

Further, claim 24 recites "wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment." *Ueno* fails to teach this further element of claim 24. As discussed above, *Ueno* fails to teach the recited segment-based model of the media server's memory, and thus further fails to teach any such segment-based model that comprises the elements recited in claim 24. Therefore, the rejection of claim 24 should also be overturned.

Independent Claim 29 and Dependent Claims 30-32

Independent claim 29 recites:

> Computer-executable software stored to a computer-readable medium, the computer-executable software comprising:
> code for <u>creating a segment-based model of a media server's memory</u>; and
> code for <u>determining whether to serve a requested streaming file from the media server based at least in part on the segment-based model</u> of the media server's memory. (Emphasis added).

As discussed above with claim 23, *Ueno* fails to teach creating a segment-based model of a media server's memory, and thus also fails to teach determining whether to serve a requested streaming file from the media server based at least in part on the segment-based model.

Accordingly, *Ueno* fails to teach all elements of claim 29, and therefore the rejection of claim 29 should be overturned.

Claims 30-32 each depend either directly or indirectly from independent claim 29, and are thus likewise believed to be allowable at least based on their dependency from claim 29 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claims 30-32 also be overturned.

Dependent Claim 33

Dependent claim 33 depends from claim 29, and thus inherits all of the limitations of claim 29 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 33 is allowable at least because of its dependence from claim 29 for the reasons discussed above.

Further, claim 33 recites "wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment." *Ueno* fails to teach this further element of claim 33. As discussed above, *Ueno* fails to teach the recited segment-based model of the media server's memory, and thus further fails to teach any such segment-based model that comprises the elements recited in claim 33. Therefore, the rejection of claim 33 should also be overturned.

Independent Claim 37 and Dependent Claim 39

Independent claim 37 recites:

> A cost-aware admission control system comprising:
> means for receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server;
> means for creating a segment-based model of the media server's memory as of time T_{cur}; and
> means for determining, based at least in part on the segment-based model

> <u>of the media server's memory, whether to accept the received request</u> for service by the media server. (Emphasis added).

As discussed above with claim 23, *Ueno* fails to teach creating a segment-based model of a media server's memory, and thus also fails to teach determining whether to accept a received request based at least in part on the segment-based model. Accordingly, *Ueno* fails to teach all elements of claim 37, and therefore the rejection of claim 37 should be overturned.

Claim 39 depends from independent claim 37, and is thus likewise believed to be allowable at least based on its dependency from claim 37 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claim 39 also be overturned.

<u>Dependent Claim 38</u>

Dependent claim 38 depends from claim 37, and thus inherits all of the limitations of claim 37 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 38 is allowable at least because of its dependence from claim 37 for the reasons discussed above.

Further, claim 38 recites "wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment." *Ueno* fails to teach this further element of claim 38. As discussed above, *Ueno* fails to teach the recited segment-based model of the media server's memory, and thus further fails to teach any such segment-based model that comprises the elements recited in claim 38. Therefore, the rejection of claim 38 should also be overturned.

C. Rejections Under 35 U.S.C. § 103(a) Over *Ueno* and *Krishnamurthy*

Claims 3-5, 9-11, 13-17, 19-22, 26-28, 35-36, and 40-42 are rejected under 35 U.S.C. § 103(a) as being unpatentable over *Ueno* in view of *Krishnamurthy*. Appellant respectfully traverses these rejections below.

The test for non-obvious subject matter is whether the differences between the subject matter and the prior art are such that the claimed subject matter as a whole would have been obvious to a person having ordinary skill in the art to which the subject matter pertains. The United States Supreme Court in Graham v. John Deere and Co., 383 U.S. 1 (1966) set forth the factual inquiries which must be considered in applying the statutory test: (1) determining of the scope and content of the prior art; (2) ascertaining the differences between the prior art and the claims at issue; and (3) resolving the level of ordinary skill in the pertinent art. As discussed further hereafter, Appellant respectfully asserts that the claims include non-obvious differences over the cited art.

Further, to properly establish obviousness, sufficient reasons that would have prompted one of ordinary skill in the art to combine/modify the prior art teachings in a manner that would arrive at the claimed invention must be present. For instance, it is "important to identify a reason that would have prompted a person of ordinary skill ... to have combined the [prior art] elements in the way the claimed invention does". *KSR Int'l Co. v. Teleflex, Inc.*, No 05-1350, 550 U.S. ______ (2007).

As discussed further below, the rejections should be withdrawn because when considering the scope and content of the applied *Ueno* and *Krishnamurthy* references there are significant differences between the applied combination and claims 3-5, 9-11, 13-17, 19-22, 26-28, 35-36, and 40-42, as the applied combination fails to disclose all elements of these claims. Thus, considering the lack of disclosure in the applied combination of all elements of claims 3-5, 9-11, 13-17, 19-22, 26-28, 35-36, and 40-42, one of ordinary skill in the art would not find these claims obvious under 35 U.S.C. §103, and therefore the rejections should be withdrawn at least for this reason.

Dependent Claim 3

Dependent claim 3 depends from claim 2, which depends from claim 1, and thus claim 3 inherits all of the limitations of claims 1 and 2 in addition to its own supplied limitations. It is

respectfully submitted that dependent claim 3 is allowable at least because of its dependence from claims 1 and 2 for the reasons discussed above.

Further, claim 3 recites "determining from the segment-based memory model a cost associated with serving the requested streaming media file from the streaming media server." The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 3. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests a segment-based memory mode, and thus neither references teaches or suggests determining from such a segment-based memory model a cost associated with serving a requested streaming media file, as recited by claim 3.

Therefore, the rejection of claim 3 should also be overturned.

<u>Dependent Claims 4 and 9</u>

Claims 4 and 9 each depends from independent claim 1, and are thus likewise believed to be allowable at least based on their dependency from claim 1 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claims 4 and 9 also be overturned.

<u>Dependent Claim 5</u>

Dependent claim 5 depends indirectly from claim 1, and thus inherits all of the limitations of claim 1 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 5 is allowable at least because of its dependence from claim 1 for the reasons discussed above.

Claim depends from claim 4, which recites "wherein said resource availability check comprises: determining a cost associated with serving the requested streaming media file from the streaming media server." Further, claim 5 recites "wherein the cost comprises: a cost of serving the requested streaming media file either from memory or from disk." The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 5. That is,

neither *Ueno* nor *Krishnamurthy* teaches or suggests determining a cost of serving the requested streaming media file either from memory or from disk, as recited by claim 5.

The Final Office Action asserts (on page 12 thereof) that *Ueno* at Col. 11, lines 10-25, Col. 8, lines 1-11, Col. 16, line 49 – col. 17, line 26, and *Krishnamurthy* at col. 5, lines 38-51 disclose the determination of a cost of serving a requested streaming media file either from memory or from disk. Below, Appellant addresses each of the cited portions of the references below, and explains that the Examiner has failed to identify any disclosure of the above-element of claim 5 in the applied references.

First, as discussed above with claim 1, col. 16, line 49 – col. 17, line 26 of *Ueno* merely describes controlling transmission rate of data from a server to a set-top unit to enable the FIFO on the set-top unit to handle the received data without overflowing. This in no way teaches determining a cost of serving the requested streaming media file either from memory or from disk.

Further, col. 11, lines 10-25 of *Ueno* provides:

> While it has been described that the source is transmitted from the output port 201b in order from the end, it is not always required to be transmitted in order from the very end in the case of a long source, but the whole source may be divided into a number of parts to perform the aforementioned transmission with respect to each part. In this case, the switching of the stream, the setting of connection and the disconnection as set forth above are repeated by the number of the division. In the aforementioned preferred embodiment, while the data stored in the storage unit has been transmitted at the same time as the transmission of the data reproduced in real time, it is considered that the data stored therein are not always transmitted at the same time so that they have been previously stored. This is the case that, for example, the storage unit is commonly used by a plurality of clients, and when a client accesses the program watched by a previous client, the data stored for the previous client have been already stored in the storage unit to be reused.

As can be clearly seen, this does not provide any teaching whatsoever of determining a cost of serving the requested streaming media file either from memory or from disk as recited by claim 5.

Further, col. 8, lines 1-11 of *Ueno* provides:

> As the quality in this case, the real-time characteristic and the low cell-discarding characteristic are required. Therefore, a class of a sufficiently guaranteed quality, such as the CBR, is used. On the other hand, at the same time as the offer of call, an information, which is not required to be offered in real time, is transmitted to the same receiving apparatus, for example, in the retrospective direction from the end of the information source, using a best effort communication class, such as the ABR and the UBR, as a second communication line, which is inexpensive although it is in non real time.

As can be clearly seen, this further portion of *Ueno* also does not provide any teaching whatsoever of determining a cost of serving the requested streaming media file either from memory or from disk as recited by claim 5. Instead, this portion of *Ueno* merely describes that different classes of communication, such as guaranteed quality class (e.g., CBR) and best effort class (e.g., ABR and UBR) can be used for communicating data.

Col. 5, lines 38-51 of *Krishnamurthy* provides:

> A primary objective of the present invention is to provide end users with a way to make dynamic network resource reservations that are suited to their financial limits and applications needs. These reservations are made on an end-to-end basis without the need to keep state information at the core routers in the network. For the network provider, pricing-based quality of service provides a simple network architecture that allows quality of service differentiation while simultaneously incorporating the admission control and congestion avoidance into the pricing schemes. Network resources are allocated dynamically when they are needed so that the network provider can allocate network resources more efficiently and provide predictable quality of service levels without over-dimensioning the network.

As can be clearly seen, this cited portion of *Krishnamurthy* also fails to provide any teaching whatsoever of determining a cost of serving the requested streaming media file either from memory or from disk as recited by claim 5. Indeed, this portion of *Krishnamurthy* does not make any mention or differentiation between serving a requested streaming media file from memory or from disk.

In view of the above, the applied combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 5. Therefore, the rejection of claim 5 should also be overturned.

Dependent Claim 10

Dependent claim 10 depends from claim 1, and thus inherits all of the limitations of claim 1 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 10 is allowable at least because of its dependence from claim 1 for the reasons discussed above.

Further, claim 10 recites "if determined that the streaming media server does not have sufficient available resources to service the new request or determined that acceptance of the new request will violate, at any point in the future, said desired quality of service provided by the streaming media server for any previously accepted requests, then rejecting the new request for service by the streaming media server." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 10. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining whether acceptance of a new request will violate, at any point in the future, a desired quality of service provided for any previously accepted requests, and thus the applied combination fails to teach or suggest this further element of claim 10. Therefore, the rejection of claim 10 should also be overturned.

Independent Claim 11 and Dependent Claims 14-15

Independent claim 11 recites:

> A method for managing admission of requests to a media server, the method comprising:
>
> receiving a new request for a streaming file to be served by a media server;
>
> determining a cost to the media server for serving the requested streaming file, wherein the cost corresponds to the media server's resources to be consumed in serving the requested streaming file, and wherein said determining said cost comprises
>
> determining a segment-based memory model that identifies content of the media server's memory as of a time that the new request is received, and

using the segment-based memory model to determine whether at least a portion of the requested streaming file is in the media server's memory; and

determining, based at least in part on the cost, whether to admit the new request for service by the media server. (Emphasis added).

The applied combination of *Ueno* and *Krishnamurthy* fails to teach or suggest at least the above-emphasized elements of claim 11. The Final Office Action (at page 13 thereof) cites to portions of *Ueno* and portions of *Krishnamurthy* as teaching determining a segment-based memory model that identifies content of the media server's memory as of a time that the new request is received. As discussed above with claim 23, *Ueno* fails to teach or suggest determining such a segment-based memory model. Further, *Krishnamurthy* fails to teach or suggest this element, as discussed below.

The Final Office Action (at page 13 thereof) cites to col. 2, lines 43-51 of *Krishnamurthy* as teaching determining a segment-based memory model that identifies content of the media server's memory as of a time that the new request is received. Col. 2, lines 43-51 of *Krishnamurthy* makes no mention of determining a memory model of a media server's memory at all, and certainly fails to teach or suggest determining a segment-based memory model as recited by claim 11. For instance, col. 2, lines 43-51 of *Krishnamurthy* merely provides:

> Additionally, the network resources are monitored and are configured to provide a plurality of predictable and dynamically variable quality of service levels, with each quality of service level guaranteeing a particular combination of network resources and including a price of service. The price of service of each quality of service level is set to optimize the admission of transmission data through the network and to avoid congestion within the network.

This makes no mention whatsoever of determining any memory model of a media server's memory. Appellant fails to see any teaching or suggestion in *Krishnamurthy* of a segment-based memory model as recited by claim 11.

In view of the above, the combination of *Ueno* and *Krishnamurthy* fails to teach or suggest all elements of claim 11, and therefore the rejection of claim 11 should be overturned.

Claims 14-15 each depends from independent claim 11, and are thus likewise believed to be allowable at least based on their dependency from claim 11 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claims 14-15 also be overturned.

Dependent Claim 13

Dependent claim 13 depends from claim 11, and thus inherits all of the limitations of claim 11 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 13 is allowable at least because of its dependence from claim 11 for the reasons discussed above.

Further, claim 13 recites "wherein the cost comprises: a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the media server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the media server's memory." The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 13. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the media server's memory, and determining a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the media server's memory. No such cost of serving the file from memory or cost of serving the file from disk is taught or suggested by *Ueno* or *Krishnamurthy*. Therefore, the rejection of claim 13 should also be overturned.

Dependent Claim 16

Dependent claim 16 depends from claim 11, and thus inherits all of the limitations of claim 11 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 16 is allowable at least because of its dependence from claim 11 for the reasons discussed above.

Further, claim 16 recites "wherein said determining whether to admit the new request for service by the media server further comprises: performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 16. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining whether acceptance of a new request will violate at any point in the future a desired quality of service provided for any previously accepted requests. As discussed above with claim 1, *Ueno* instead merely evaluates whether a pre-set resource utilization threshold is currently met when determining whether to accept a new request, without any consideration of whether such acceptance would violate at any point in the future a desired quality of service for any previously accepted requests. Likewise, *Krishnamurthy* also fails to teach or suggest this element. Therefore, the rejection of claim 16 should also be overturned.

Independent Claim 17 and Dependent Claims 20-21

Independent claim 17 recites:

> A system comprising:
> server having a memory, wherein said server is operable to serve at least one streaming file to clients communicatively coupled thereto; and
> an admission controller operable to receive a new request for a streaming file to be served by said server, determine a cost to the server for serving the requested streaming file, wherein the cost corresponds to the server's resources to be consumed in serving the requested streaming file, and determine, based at least in part on the cost, whether to admit the new request for service by the server;
> wherein said admission controller is further operable to determine a segment-based memory model that identifies content of the server's memory as of a time that the new request is received, and said admission controller is operable to use the segment-based memory model to determine whether at least a portion of the requested streaming file is in the server's memory. (Emphasis added).

The applied combination of *Ueno* and *Krishnamurthy* fails to teach or suggest at least the above-emphasized elements of claim 17. In treating claim 18, the Final Office Action (at page 15 thereof) cites to portions of *Ueno* and portions of *Krishnamurthy* as teaching determining a

segment-based memory model that identifies content of the server's memory as of a time that the new request is received. As discussed above with claim 23, *Ueno* fails to teach or suggest determining such a segment-based memory model. Further, *Krishnamurthy* fails to teach or suggest this element, as discussed below.

The Final Office Action (at page 15 thereof) cites to col. 2, lines 43-51 of *Krishnamurthy* as teaching determining a segment-based memory model that identifies content of the server's memory as of a time that the new request is received. As discussed above with claim 11, *Krishnamurthy* fails to teach or suggest determining such a memory model.

In view of the above, the combination of *Ueno* and *Krishnamurthy* fails to teach or suggest all elements of claim 17, and therefore the rejection of claim 17 should be overturned.

Claims 20-21 each depends from independent claim 17, and are thus likewise believed to be allowable at least based on their dependency from claim 17 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claims 20-21 also be overturned.

Dependent Claim 19

Dependent claim 19 depends from claim 17, and thus inherits all of the limitations of claim 17 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 19 is allowable at least because of its dependence from claim 17 for the reasons discussed above.

Further, claim 19 recites "wherein the cost comprises: a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the server's memory." The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 19. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the media server's memory,

and determining a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the media server's memory. No such cost of serving the file from memory or cost of serving the file from disk is taught or suggested by *Ueno* or *Krishnamurthy*. Therefore, the rejection of claim 19 should also be overturned.

Dependent Claim 22

Dependent claim 22 depends from claim 17, and thus inherits all of the limitations of claim 17 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 22 is allowable at least because of its dependence from claim 17 for the reasons discussed above.

Further, claim 22 recites "wherein said admission controller is further operable to perform quality of service guarantee check for the server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the server for any previously accepted requests." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 22. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining whether acceptance of a new request will violate at any point in the future a desired quality of service provided for any previously accepted requests. As discussed above with claim 1, *Ueno* instead merely evaluates whether a pre-set resource utilization threshold is currently met when determining whether to accept a new request, without any consideration of whether such acceptance would violate at any point in the future a desired quality of service for any previously accepted requests. Likewise, *Krishnamurthy* also fails to teach or suggest this element. Therefore, the rejection of claim 22 should also be overturned.

Dependent Claims 26-27

Claims 26-27 each depends from independent claim 23, and are thus likewise believed to be allowable at least based on their dependency from claim 23 for the reasons discussed above.

Accordingly, Appellant respectfully requests that the rejection of claims 26-27 also be overturned.

Dependent Claim 28

Dependent claim 28 depends from claim 23, and thus inherits all of the limitations of claim 23 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 28 is allowable at least because of its dependence from claim 23 for the reasons discussed above.

Further, claim 28 recites "wherein said determining whether to accept the received request for service by the media server comprises: performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 28. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining whether acceptance of a new request will violate at any point in the future a desired quality of service provided for any previously accepted requests. As discussed above with claim 1, *Ueno* instead merely evaluates whether a pre-set resource utilization threshold is currently met when determining whether to accept a new request, without any consideration of whether such acceptance would violate at any point in the future a desired quality of service for any previously accepted requests. Likewise, *Krishnamurthy* also fails to teach or suggest this element. Therefore, the rejection of claim 28 should also be overturned.

Dependent Claim 35

Claim 35 depends from independent claim 29, and is thus likewise believed to be allowable at least based on its dependency from claim 29 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claim 35 also be overturned.

Dependent Claim 36

Dependent claim 36 depends from claim 29, and thus inherits all of the limitations of claim 29 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 36 is allowable at least because of its dependence from claim 29 for the reasons discussed above.

Further, claim 36 recites "wherein said code for determining whether to serve a requested streaming file from the media server comprises: code for performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 36. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining whether acceptance of a new request will violate at any point in the future a desired quality of service provided for any previously accepted requests. As discussed above with claim 1, *Ueno* instead merely evaluates whether a pre-set resource utilization threshold is currently met when determining whether to accept a new request, without any consideration of whether such acceptance would violate at any point in the future a desired quality of service for any previously accepted requests. Likewise, *Krishnamurthy* also fails to teach or suggest this element. Therefore, the rejection of claim 36 should also be overturned.

Dependent Claim 40

Claim 40 depends from independent claim 37, and is thus likewise believed to be allowable at least based on its dependency from claim 37 for the reasons discussed above. Accordingly, Appellant respectfully requests that the rejection of claim 40 also be overturned.

Dependent Claim 41

Dependent claim 41 depends from claim 1, and thus inherits all of the limitations of claim 1 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 41 is allowable at least because of its dependence from claim 1 for the reasons discussed above.

As discussed above, claim 1 recites "performing a quality of service guarantee check for the streaming media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests" (emphasis added), which Appellant maintains is not taught or suggested by the applied references. Further, claim 41 depends from claim 1 and further recites "wherein said performing said quality of service guarantee check comprises: performing said quality of service guarantee check even when determined by said resource availability check that the streaming media server has sufficient available resources to service the new request." (Emphasis added). Thus, claim 41 recites that the quality of service guarantee check (to determine whether acceptance of the new request will violate at any point in the future a desired quality of service provided for any previously accepted requests), even when determined by said resource availability check that the streaming media server has sufficient available resources to service the new request. The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 41. Rather, the applied references appear at most to disclose determining whether sufficient resource availability exists for servicing a newly received request, and does not teach or suggest performing the above-mentioned quality of service guarantee check when determined that the serve has sufficient available resources to service the new request. As discussed above with claim 1, *Ueno* instead merely evaluates whether a pre-set resource utilization threshold is currently met when determining whether to accept a new request, and if the utilization threshold is not met (i.e., sufficient resources exist to service the new request), the request is accepted without performing any further quality of service guarantee check with regard to any previously accepted requests. Likewise, *Krishnamurthy* also fails to teach or suggest this element. Therefore, the rejection of claim 41 should also be overturned.

<u>Dependent Claim 42</u>

Dependent claim 42 depends from claim 17, and thus inherits all of the limitations of claim 17 in addition to its own supplied limitations. It is respectfully submitted that dependent claim 42 is allowable at least because of its dependence from claim 17 for the reasons discussed above.

Further, claim 42 recites "wherein said admission controller determines the cost to the server for serving the requested streaming file <u>based at least in part on the determined segment-based memory model</u>." (Emphasis added). The combination of *Ueno* and *Krishnamurthy* fails to teach or suggest this further element of claim 42. That is, neither *Ueno* nor *Krishnamurthy* teaches or suggests determining a cost for serving the requested streaming file based at least in part on the determined segment-based memory model. Therefore, the rejection of claim 42 should also be overturned.

Conclusion

In view of the above, Appellant requests that the board overturn the outstanding rejections of claims 1-11, 13-17 and 19-42. Attached hereto are a Claims Appendix, Evidence Appendix, and Related Proceedings Appendix. As noted in the attached Evidence Appendix, no evidence pursuant to §§ 1.130, 1.131, or 1.132 or entered by or relied upon by the Examiner is being submitted. Also, as noted by the Related Proceedings Appendix, no related proceedings are referenced in II above, and thus no copies of decisions in related proceedings are provided.

Applicant believes a fee of $500.00 is due with this response. However, if any additional fee is due, please charge our Deposit Account No. 08-2025, under Order No. 200311046-1 from which the undersigned is authorized to draw.

Respectfully submitted,

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office electronic filing system in accordance with § 1.6(a)(4).

Dated: July 19, 2007

Signature: Donna Forbit
(Donna Forbit)

By Jody C. Bishop
Jody C. Bishop
Registration No.: 44,034
Attorney/ Agent for Applicant(s)
Dated: July 19, 2007
Telephone No. (214) 855-8007

VIII. CLAIMS APPENDIX

Claims Involved in the Appeal of Application Serial No. 10/601,992

1. A method for managing admission of requests to a streaming media server, the method comprising:

receiving a new request for a streaming media file to be served by a streaming media server;

performing a resource availability check for the streaming media server to determine whether the streaming media server has sufficient available resources to service the new request; and

performing a quality of service guarantee check for the streaming media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the streaming media server for any previously accepted requests.

2. The method of claim 1 wherein said resource availability check comprises:

using a segment-based memory model to determine whether at least a portion of the requested streaming media file is in the streaming media server's memory.

3. The method of claim 2 further comprising:

determining from the segment-based memory model a cost associated with serving the requested streaming media file from the streaming media server.

4. The method of claim 1 wherein said resource availability check comprises:

determining a cost associated with serving the requested streaming media file from the streaming media server.

5. The method of claim 4 wherein the cost comprises:

a cost of serving the requested streaming media file either from memory or from disk.

6. The method of claim 1 wherein said resources comprise memory resources and disk resources.

7. The method of claim 1 wherein said sufficient available resources to service the new request comprises sufficient resources available so as not to overload the streaming media server.

8. The method of claim 1 wherein said desired quality of service comprises real-time delivery of streaming media files requested by said previously accepted requests.

9. The method of claim 1 further comprising:

if determined that the streaming media server has sufficient available resources to service the new request and determined that acceptance of the new request will not violate, at any point in the future, said desired quality of service provided by the streaming media server for any previously accepted requests, then the streaming media server serving the requested streaming media file for said new request.

10. The method of claim 1 further comprising:

if determined that the streaming media server does not have sufficient available resources to service the new request or determined that acceptance of the new request will violate, at any point in the future, said desired quality of service provided by the streaming media server for any previously accepted requests, then rejecting the new request for service by the streaming media server.

11. A method for managing admission of requests to a media server, the method comprising:

receiving a new request for a streaming file to be served by a media server;

determining a cost to the media server for serving the requested streaming file, wherein the cost corresponds to the media server's resources to be consumed in serving the requested streaming file, and wherein said determining said cost comprises

determining a segment-based memory model that identifies content of the media server's memory as of a time that the new request is received, and

using the segment-based memory model to determine whether at least a portion of the requested streaming file is in the media server's memory; and

determining, based at least in part on the cost, whether to admit the new request for service by the media server.

12. (Canceled)

13. The method of claim 11 wherein the cost comprises:

a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the media server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the media server's memory.

14. The method of claim 11 wherein said determining whether to admit the new request for service by the media server comprises:

performing a resource availability check for the media server to determine whether the media server has sufficient available resources to service the new request.

15. The method of claim 14 wherein said sufficient available resources to service the new request comprises sufficient resources available so as not to overload the media server.

16. The method of claim 14 wherein said determining whether to admit the new request for service by the media server further comprises:

performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests.

17. A system comprising:

server having a memory, wherein said server is operable to serve at least one streaming file to clients communicatively coupled thereto; and

an admission controller operable to receive a new request for a streaming file to be served by said server, determine a cost to the server for serving the requested streaming file, wherein the cost corresponds to the server's resources to be consumed in serving the requested streaming file, and determine, based at least in part on the cost, whether to admit the new request for service by the server;

wherein said admission controller is further operable to determine a segment-based memory model that identifies content of the server's memory as of a time that the new request is received, and said admission controller is operable to use the segment-based memory model to determine whether at least a portion of the requested streaming file is in the server's memory.

18. (Canceled)

19. The system of claim 17 wherein the cost comprises:

a cost of serving the requested streaming file from memory if determined that the requested streaming file is in the server's memory and a cost of serving the requested streaming file from disk if determined that the requested streaming file is not in the server's memory.

20. The system of claim 17 wherein said admission controller is further operable to perform a resource availability check for the server to determine whether the server has sufficient available resources to service the new request.

21. The system of claim 20 wherein said sufficient available resources to service the new request comprises sufficient resources available so as not to overload the server.

22. The system of claim 17 wherein said admission controller is further operable to perform quality of service guarantee check for the server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the server for any previously accepted requests.

23. A method comprising:

receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server;

creating a segment-based model of the media server's memory as of time T_{cur}; and

based at least in part on the segment-based model of the media server's memory, determining whether to accept the received request for service by the media server.

24. The method of claim 23 wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment.

25. The method of claim 23 wherein said determining whether to accept the received request for service by the media server comprises:

determining whether the received request can be serviced by the media server without overloading the media server.

26. The method of claim 23 wherein said determining whether to accept the received request for service by the media server comprises:

determining a cost to the server for serving the requested streaming file, wherein the cost corresponds to the amount of the media server's resources to be consumed in serving the requested streaming file.

27. The method of claim 23 wherein said determining whether to accept the received request for service by the media server comprises:

performing a resource availability check for the media server to determine whether the media server has sufficient available resources to service the new request.

28. The method of claim 23 wherein said determining whether to accept the received request for service by the media server comprises:

performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests.

29. Computer-executable software stored to a computer-readable medium, the computer-executable software comprising:

code for creating a segment-based model of a media server's memory; and

code for determining whether to serve a requested streaming file from the media server based at least in part on the segment-based model of the media server's memory.

30. The computer-executable software code of claim 29 further comprising:

code for receiving a request for said streaming file.

31. The computer-executable software code of claim 30 further comprising:

code, responsive to receiving said request, for determining whether to accept the request for service by the media server.

32. The computer-executable software code of claim 31 wherein said code for determining whether to accept the request for service by the media server comprises:

code for determining whether the request can be serviced by the media server without overloading the media server.

33. The computer-executable software code of claim 29 wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment.

34. The computer-executable software code of claim 29 wherein said code for determining whether to serve a requested streaming file from the media server comprises:

code for determining a cost to the media server for serving the requested streaming file, wherein the cost corresponds to the amount of the media server's resources to be consumed in serving the requested streaming file.

35. The computer-executable software of claim 29 wherein said code for determining whether to serve a requested streaming file from the media server comprises:

code for performing a resource availability check for the media server to determine whether the media server has sufficient available resources to service the new request.

36. The computer-executable software code of claim 29 wherein said code for determining whether to serve a requested streaming file from the media server comprises:

code for performing quality of service guarantee check for the media server to determine whether acceptance of the new request will violate, at any point in the future, a desired quality of service provided by the media server for any previously accepted requests.

37. A cost-aware admission control system comprising:

means for receiving, at a time T_{cur}, a new request for a streaming file to be served by a media server;

means for creating a segment-based model of the media server's memory as of time T_{cur}; and

means for determining, based at least in part on the segment-based model of the media server's memory, whether to accept the received request for service by the media server.

38. The cost-aware admission control system of claim 37 wherein said segment-based model of the media server's memory comprises (a) identification of unique segments of streaming files previously accessed by clients of the media server and (b) identification of corresponding timestamps of most recent accesses of each unique segment.

39. The cost-aware admission control system of claim 37 wherein said means for determining whether to accept the received request for service by the media server comprises:

means for determining whether the received request can be serviced by the media server without overloading the media server.

40. The cost-aware admission control system of claim 37 wherein said means for determining whether to accept the received request for service by the media server comprises:

means for determining a cost to the server for serving the requested streaming file, wherein the cost corresponds to the amount of the media server's resources to be consumed in serving the requested streaming file.

41. The method of claim 1 wherein said performing said quality of service guarantee check comprises:

performing said quality of service guarantee check even when determined by said resource availability check that the streaming media server has sufficient available resources to service the new request.

42. The system of claim 17 wherein said admission controller determines the cost to the server for serving the requested streaming file based at least in part on the determined segment-based memory model.

IX. EVIDENCE APPENDIX

No evidence pursuant to §§ 1.130, 1.131, or 1.132 or entered by or relied upon by the examiner is being submitted.

X. RELATED PROCEEDINGS APPENDIX

No related proceedings are referenced in II above, and thus no copies of decisions in related proceedings are provided.